Predicting housing Sale Price

Based on Ames Iowa Housing Dataset (2006-2010)

Problem Statement

Using Ames Housing Data obtained directly from from the Ames Assessor’s Office year 2006 to 2010, to build a model that best predict the housing prices based on the given 80 housing features to select from and model with.

Findings and results from this model would be used to advise property agents on how they can best maximize profits.

Data Cleaning / Feature Engineering

Data Cleaning

```
In [7]: print('Train set shape: ', df_train.shape)
        print('Test set shape: ', df_test.shape)
        print('Columns that are different between 2 dataset: ', df_train.columns.difference(df_test.columns))
```

```
Train set shape:  (2051, 81)
Test set shape:  (879, 80)
Columns that are different between 2 dataset:  Index(['SalePrice'], dtype='object')
```

- **Concatenate both datasets into 1 for data cleaning**
- **To minimize chances of different variables, sequences, etc.**
- **SalePrice is the only different variable - easy for splitting back to Train / Test set after cleaning**

Data Cleaning

```
In [7]: # showing all the variables that have missing data

        null_counts = df_combined.isnull().sum()
        null_counts[null_counts > 0]
```

Out[7]:

Lot Frontage	490
Alley	2732
Mas Vnr Type	23
Mas Vnr Area	23
Bsmt Qual	80
Bsmt Cond	80
Bsmt Exposure	83
BsmtFin Type 1	80
BsmtFin SF 1	1
BsmtFin Type 2	81
BsmtFin SF 2	1
Bsmt Unf SF	1
Total Bsmt SF	1
Electrical	1
Bsmt Full Bath	2
Bsmt Half Bath	2
Fireplace Qu	1422
Garage Type	157
Garage Yr Blt	159
Garage Finish	159
Garage Cars	1
Garage Area	1
Garage Qual	159
Garage Cond	159
Pool QC	2917
Fence	2358
Misc Feature	2824
SalePrice	879

dtype: int64

1. Missing data

- Lot Frontage (490)
- Groupby: MS SubClass (median)

Data Dictionary:
Lot Frontage: Linear feet of street connected to property (490)
MS SubClass : The type of dwelling involved in the sale

Data Cleaning

1. Missing data
- NA/None is an options
- NMAR

Bsmt Qual (Ordinal): Evaluates the height of the basement

Ex	Excellent (100+ inches)
Gd	Good (90-99 inches)
TA	Typical (80-89 inches)
Fa	Fair (70-79 inches)
Po	Poor (<70 inches
NA	No Basement

Bsmt Cond (Ordinal): Evaluates the general condition of the basement

Ex	Excellent
Gd	Good
TA	Typical - slight dampness allowed
Fa	Fair - dampness or some cracking or settling
Po	Poor - Severe cracking, settling, or wetness
NA	No Basement

Bsmt Exposure (Ordinal): Refers to walkout or garden level walls

Gd	Good Exposure
Av	Average Exposure (split levels or foyers typically score average or above)
Mn	Mimimum Exposure
No	No Exposure
NA	No Basement

BsmtFin Type 1 (Ordinal): Rating of basement finished area

GLQ	Good Living Quarters
ALQ	Average Living Quarters
BLQ	Below Average Living Quarters
Rec	Average Rec Room
LwQ	Low Quality
Unf	Unfinshed
NA	No Basement

Mas Vnr Type (Nominal): Masonry veneer type

BrkCmn	Brick Common
BrkFace	Brick Face
CBlock	Cinder Block
None	None
Stone	Stone

Alley (Nominal): Type of alley access to property

Grvl	Gravel
Pave	Paved
NA	No alley access

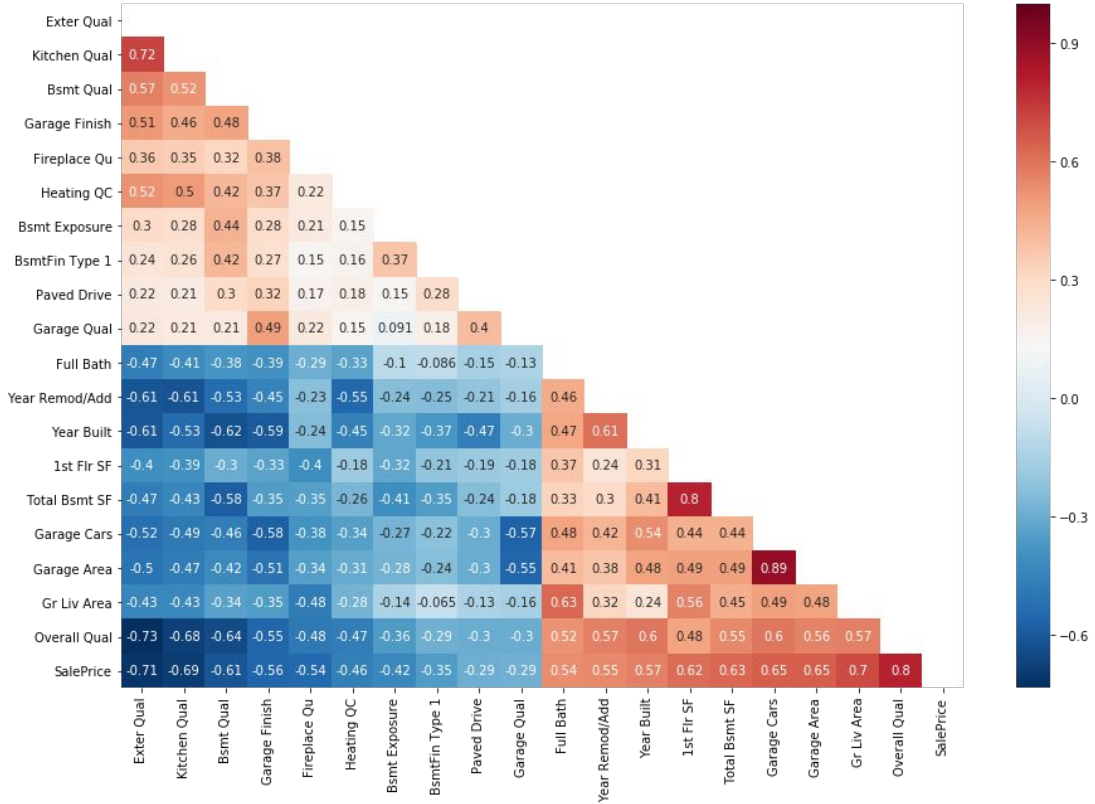
Data Cleaning

2. **Categorical (ordinal)**
   - Changed them to numeric
   - Ranking in order

Bsmt Cond (Ordinal): Evaluates the general condition of the basement

1	Ex	Excellent
2	Gd	Good
3	TA	Typical - slight dampness allowed
4	Fa	Fair - dampness or some cracking or settling
5	Po	Poor - Severe cracking, settling, or wetness
6	NA	No Basement

Data Cleaning



Top & bottom 10 numeric columns

High correlation with SalePrice (dependent variable)

Remove multicollinearity

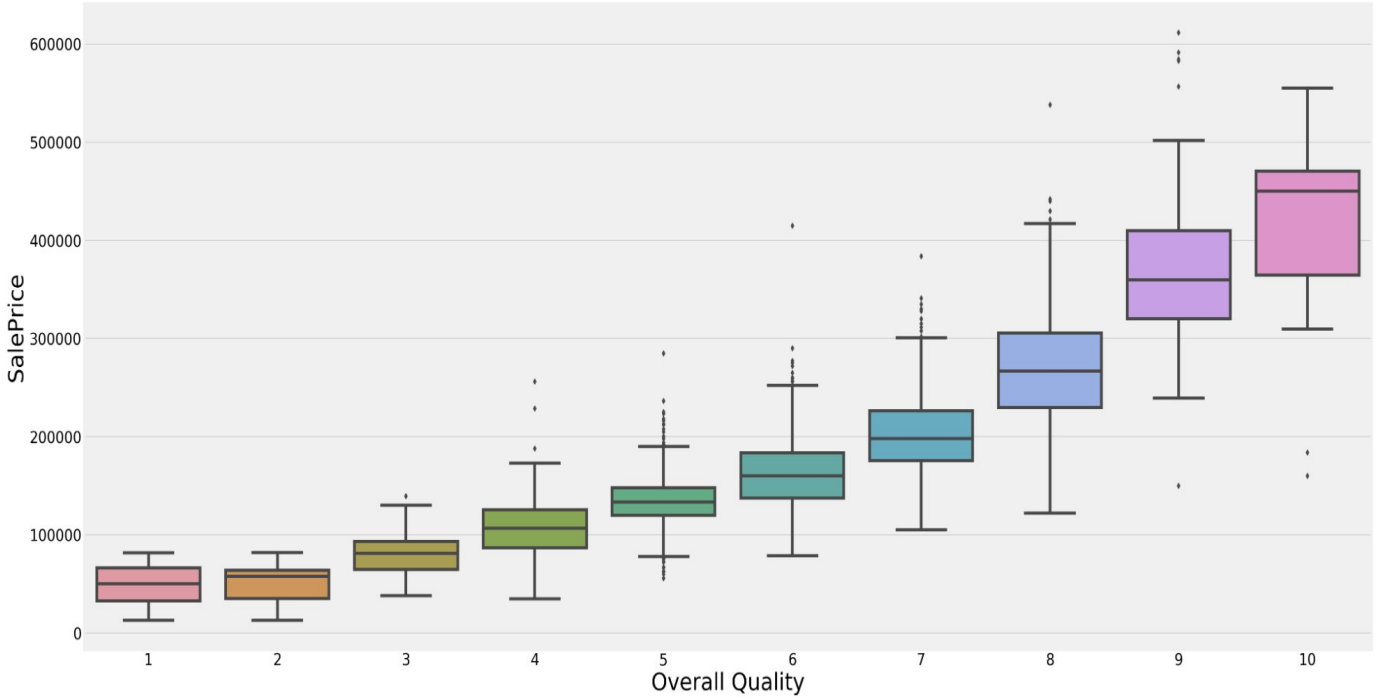
Create feature engineering

Categorical columns – dummies

Exploratory Data Analysis

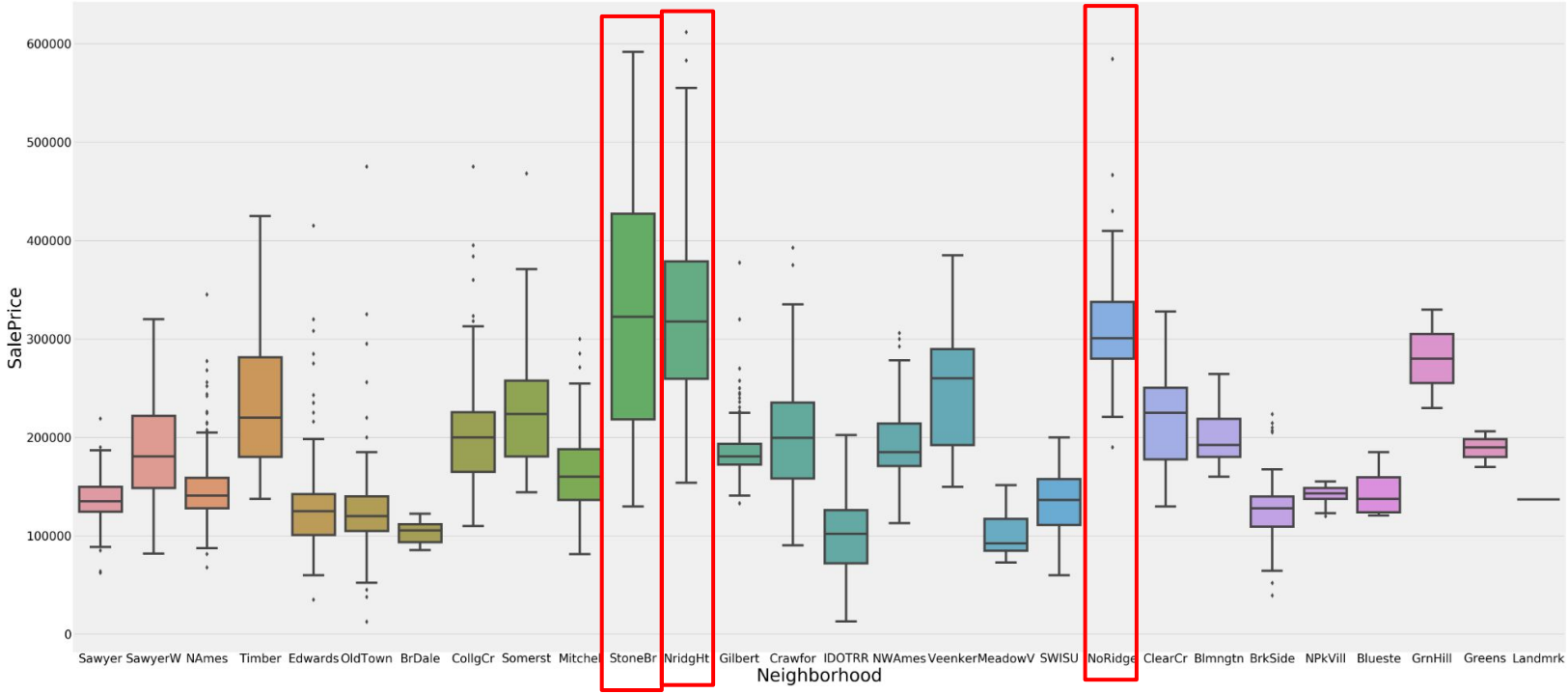
Exploratory Data Analysis

Clear association between overall quality and housing sale price



Exploratory Data Analysis

Stone Brook, Northridge Heights, and Northridge = top contenders for SalePrice



Modelling

Linear Regression
Performs worse than baseline

Ridge Regularization
Explains ~86% of variance in SalePrice

Lasso Regularization
Explains ~86% of variance in SalePrice

Feature Selection

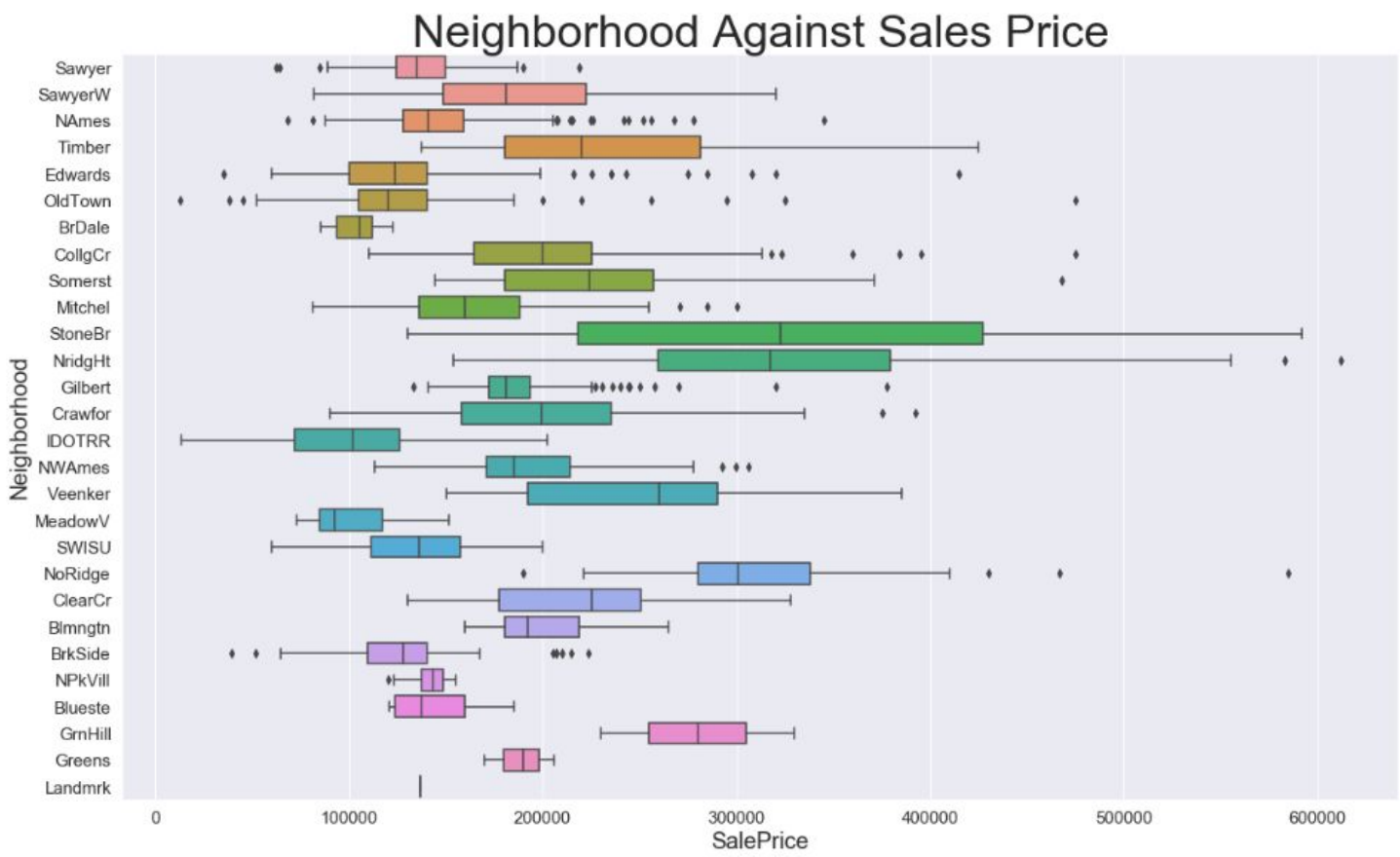
ElasticNet Regularization
Explains ~86% of variance in SalePrice

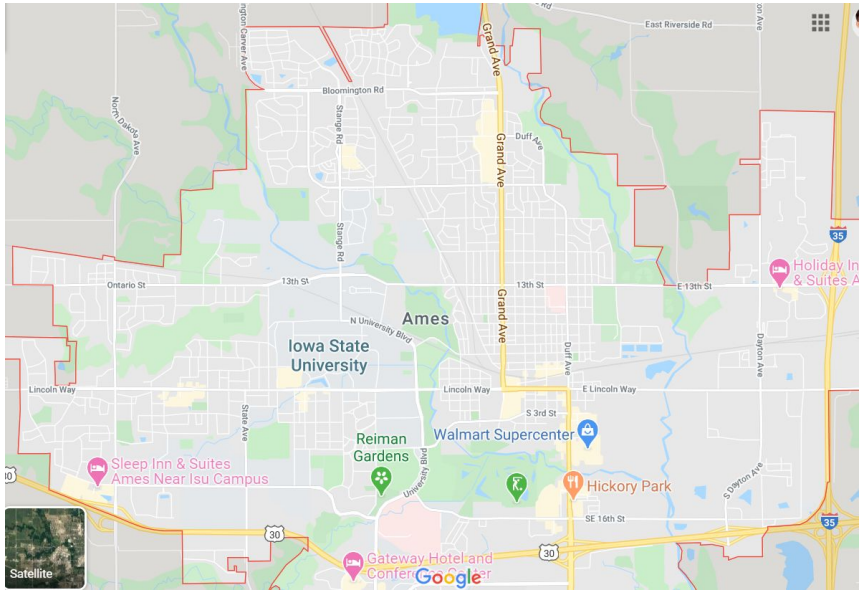
What strongly affects Sale Price?

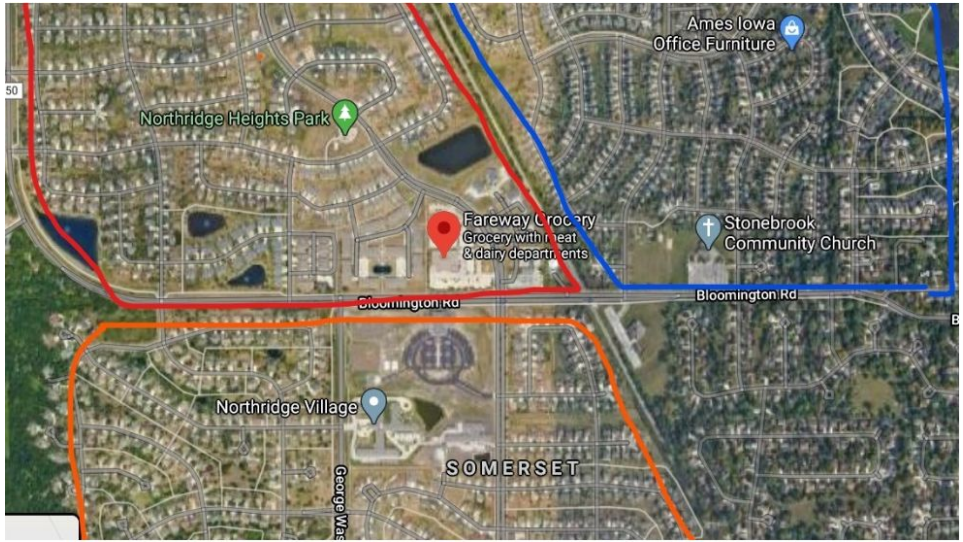
Positive predictors
Overall Quality of House
Stone Brook, Northridge Heights, Northridge neighborhoods
Overall Condition of House

Negative predictors
Town Houses
Edwards neighborhood
Stucco material for house exterior

Recommendations


Neighborhood Against Sales Price
Sawyer
SawyerW
NAmes
Timber
Edwards
OldTown
BrDale
CollgCr
Somerst
Mitchel
StoneBr
NridgHt
Gilbert
Crawfor
IDOTRR
NWAmes
Veenker
MeadowV
SWISU
NoRidge
ClearCr
Blmngtn
BrkSide
NPkVill
Blueste
GrnHill
Greens
Landmrk
Neighborhood
0
100000
200000
300000
400000
500000
600000
SalePrice

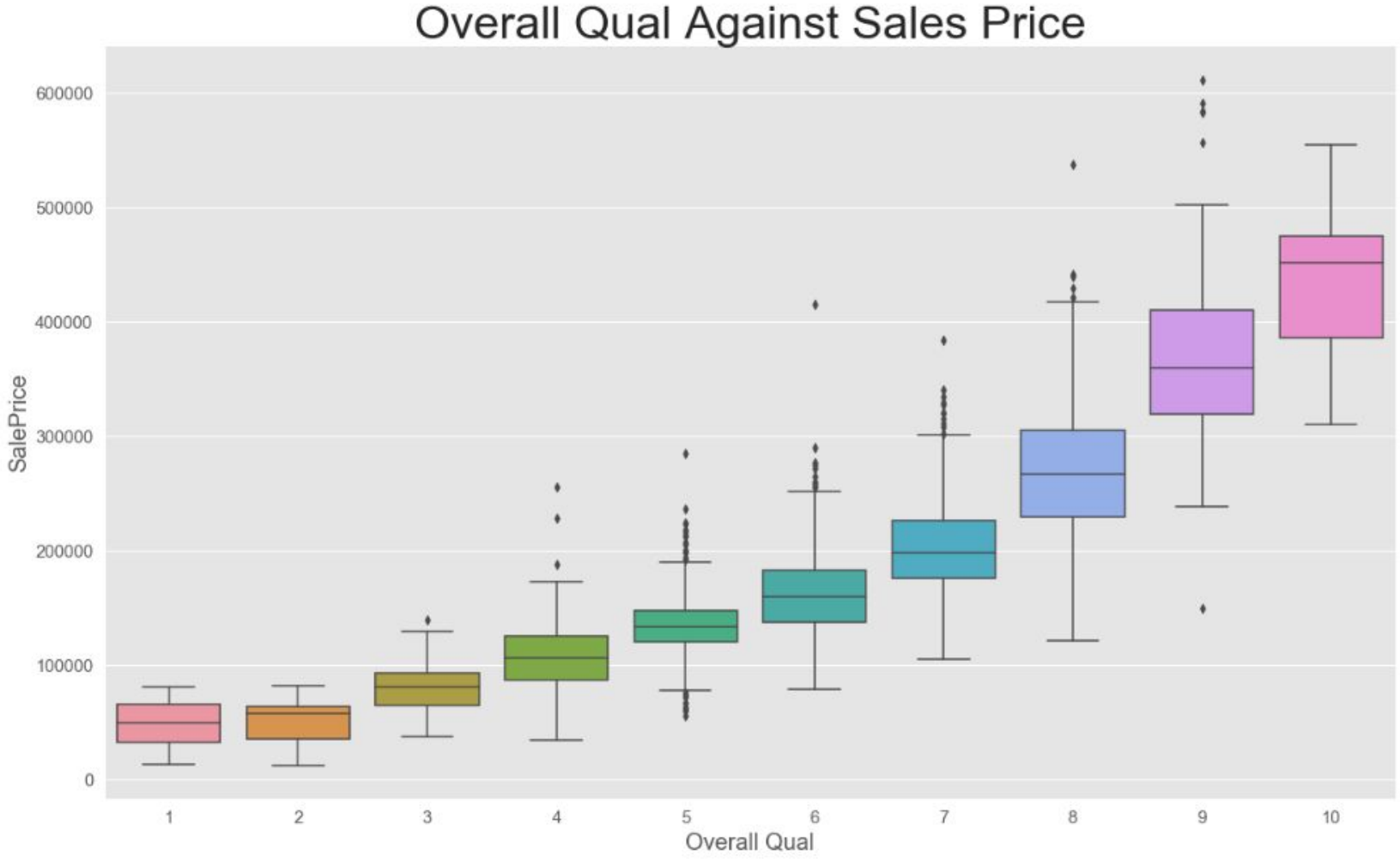

East Riverside Rd
Bloomington Rd
Grand Ave
Duff Ave
Stange Rd
Ontario St
13th St
E 13th St
Holiday In
& Suites A
Ames
N University Blvd
Iowa State
University
Lincoln Way
E Lincoln Way
S 3rd St
Reiman
Gardens
Walmart Supercenter
Sleep Inn & Suites
Ames Near Isu Campus
Hickory Park
SE 16th St
Dayton Ave
State Ave
University Blvd
30
35
Gateway Hotel and
Satellite
Google


Ames Iowa
Office Furniture
Northridge Heights Park
Fareway Grocery
Grocery with meat
& dairy departments
Stonebrook
Community Church
Bloomington Rd
Bloomington Rd
Northridge Village
SOMERSET


4014 Hastings Ave
Ames, Iowa
Google
Street View
Google


1900 Wyngate Dr
Ames, Iowa
Google
Street View
Google


3300 Valley View Rd
Ames, Iowa
Google
Street View
Google


Overall Qual Against Sales Price
SalePrice
600000
500000
400000
300000
200000
100000
0
1
2
3
4
5
6
7
8
9
10
Overall Qual

- Property Agents

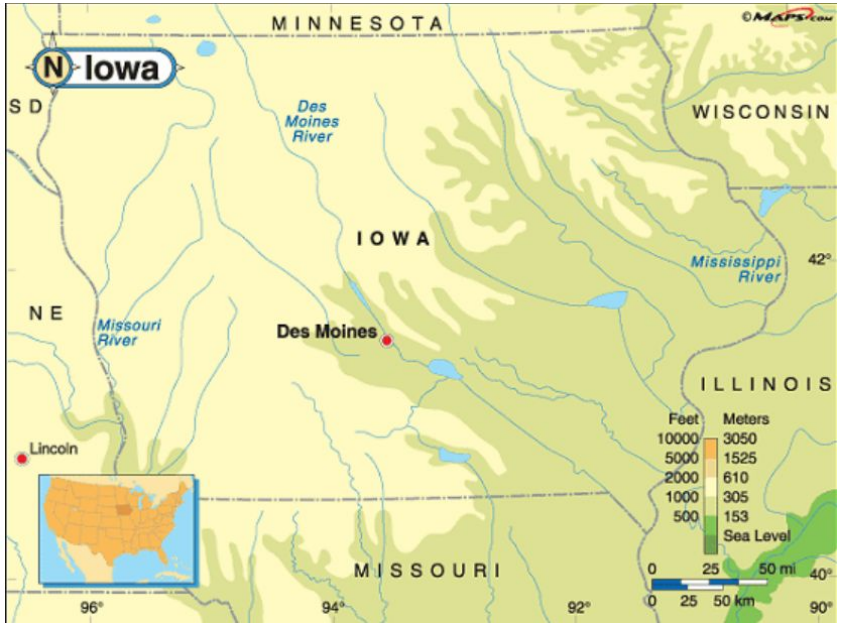
  To maximize profits, they can target areas like North Ridge Height, Stone Brooke, North Ridge which will attract higher income buyers due to sufficient amenities surrounding the vicinities like grocery shop and several schools down the road. Low crime rates within these areas will also be an appeal to them.

- Advising for Home Sellers

  Both House Quality and Year Remodel are the top 2 factors for higher sales prices. For Homeowners looking to sell, they can focus on making certain home improvements before selling their houses to improve sale prices.

- **Advising For Home Buyers**

  For Home buyers, they can look to purchase house in neighbourhoods with good amenities nearby and also with lower crime rates.
  Also avoid houses that are on lower grounds as over the years we can see that climate change had contributed to increasing flood occurrences in Iowa.



At least three people in **Iowa** and Nebraska have died. Nearly 14 million people in the midwestern and southern states have been affected by the **flooding**, which the New York Times has called "The Great **Flood** of **2019**". New record river levels were set in 42 different locations.

Property damage: $2.9 billion (1.6B in Iowa; 1....

Date: March 2019 – present

Location: Midwestern United States

en.wikipedia.org › wiki › 2019_Midwestern_U.S._floods

2019 Midwestern U.S. floods - Wikipedia

Recommendations for Further Research

For Real Estate Agents to conduct market research on population demographics if there are more younger single adults compared to family in Ames (Iowa) from sites.

Gather information on economical data e.g. salary,retrenchment rate (which may affect sales prices). Get additional housing data for Ames (Iowa) e.g. seller/buyer data, house price index and price per sqft.

https://www.census.gov/quickfacts/fact/table/IA#

Preliminary Iowa Flood Maps

New and Preliminary Iowa Flood Maps provide the public an early look at a home or community's projected risk to flood hazards. This page is for homeowners who want to understand how their current effective Flood Map may change when the preliminary FEMA maps becomes effective.

Releasing Preliminary Iowa Flood Maps

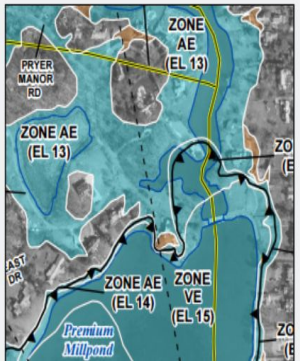
The release of preliminary flood hazard maps, or Flood Insurance Rate Maps (FIRMs), is an important step in the mapping lifecycle for a community. This release provides community officials, the public, and other stakeholders with their first view of the current flood hazards, which include changes that may have occurred in the flood risks throughout the community, or county, since the last flood hazard map was published.

Property Owners Can Take Advantage of "Grandfathering"

If a property is mapped into a high-risk area (shown as a zone labeled with letters starting with "A" or "V") and the owner has a mortgage through a federally regulated or insured lender, flood insurance will be required when the FIRM becomes effective. Lenders do have the option to make the purchase of flood insurance a condition for their loans at any time, and some lenders may institute such requirements in advance of the maps becoming effective.

FEMA
Official Flood Maps

Current Flood Maps



More research can be done on which property is located on a higher ground level as this might be a strong factor for higher sales prices.

Source : https://www.floodfind.com/iowa-flood-maps/?gclid=CjwKCAiA7t3yBRADEiwA4GFII2Z0UxFwVFotbzXLqjc049y11mZr1vJIApWwLAb6MPyrf8ekI3DS7xoCSn4QAvD_BwE